

161 Unterstützung wissenschaftlicher Datenverarbeitung mit Großrechnern durch Minirechner.

Grimm, J. (Zürich)

Departement für Innere Medizin, Universitätsspital Zürich

Im Laufe wissenschaftlicher Auswertungen auf Großrechnern muß für die organisatorischen bzw. computertechnischen Belange meist mehr Zeit aufgewendet werden als für die eigentliche statistische Programmierung. Dadurch vermindert sich für den praktischen Gebrauch die Attraktivität und Leistung großer Programmpakete wie SPSS, SAS oder BMDP beträchtlich. Im folgenden werden Programme beschrieben, die auf verschiedenen Rechnern dazu beitragen, dem Auswerter einen Teil der Routinearbeiten zu erleichtern.

Verbindungen Unispital – Rechenzentrum

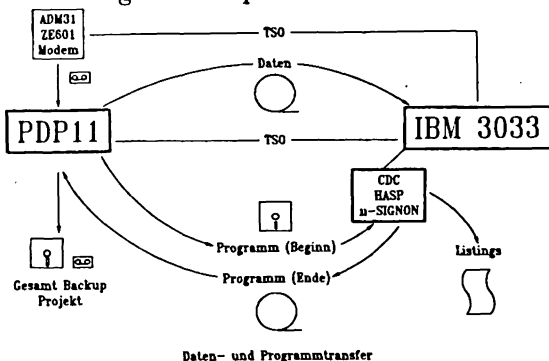


Abb. 1 Hardwarekonfiguration, die für statistische Auswertungen benutzt wird. Die Daten werden an autonomen Datenstationen, bestehend aus Terminal Lear Siegler ADM 31, Kassettenrecorder Zirelco ZE 601 (2400 Baud), sowie Akkustikmodem eingegeben. Der PDP11 Computer wird in sehr verschiedener Weise für die Daten- und Programm-Bearbeitung sowie, falls nötig, direkt an der IBM mittels Time Sharing Option (TSO) eingesetzt. Innerhalb des Spitals steht als Satellit eine HASP Workstation mit einem Control Data Cyber 18-20 und Mikroprozessor Dialogeinheit μ -SIGNON zur Verfügung. Der Einsatz der Programme ist jedoch bereits schon nur mit einem PDP11 mit Modem möglich.

Die Daten werden an autonomen, mobilen Datenstationen von EDV-fremdem Personal am gewohnten Arbeitsort eingegeben. Die Terminals sind sehr einfach zu bedienen, da sie Funktionstasten für die Cursorsteuerung sowie für das Einfügen bzw. Löschen von Zeichen oder Zeilen aufweisen. Verbunden mit den bekannten und anschaulichen Funktionen eines Kassettenrecorders wirkt sich dies nicht nur in einer schnellen Akzeptierung der Datenstationen aus, sondern auch in einem problemlosen Übergang zur späteren Arbeit am Computer selbst, bei der Korrektur der Daten. Gleichzeitig wird der Rechner durch Einsatz von Datenstationen von den oft langwierigen Daten- oder Texteingaben nicht belastet.

Die auf der Kassette enthaltenen Daten werden seriell in den PDP11 übertragen. Da dessen Systemroutinen bei einem zwar asynchronen aber kontinuierlichen Input Zeichen verlieren, wurde das Einleseprogramm DLZE in Maschinensprache erstellt, das infolge eigener Interrupt-Routinen sowie Doppelpufferung keine Zeichen verliert und gleichzeitig eine Vorprüfung der Daten erlaubt. Alle übrigen Programme wurden in Fortran unter dem Digital Equipment Betriebssystem RT 11 geschrieben.

Die Dialogeinheit μ -SIGNON am Satelliten erwies sich als notwendig, um den Rechnerdialog für ungeübtes Personal fehlerfrei zu ermöglichen. Sie wurde mit einem Intel Mikroprozessor 8748 realisiert, der mittels eines Cross-Assemblers auf dem PDP11 programmiert wurde.

Die eingelesenen Rohdaten werden in das Parameterfile mit allen Variablenamen, das Deskriptorfile, sowie schließlich in die eigentlichen Daten selbst zerlegt. Das Deskriptorfile enthält zu jeder Variablen das Minimum, Maximum und den für fehlende Meßwerte vereinbarten "Missing Value".

Das Programm ZEPDP erzeugt anhand dieses Deskriptorfiles automatisch ein File mit den Fortran-Formaten der Daten. Damit können die - bis zu diesem Zeitpunkt formatfreien Daten - formatiert und zusammen mit Fehlermeldungen bezüglich Zeichengültigkeit, Variablenzahl und Wertebereich ausgedruckt werden. Nach dem manuellen Vergleich mit den Originaldaten dient das Programm CORRECT zur Korrektur. Hierbei können Zahlen jedoch nur im Rahmen der gegebenen Fortran-Formate geändert werden, d.h. Dezimalpunkte und Anzahl Stellen sind nicht mehr zu beeinflussen. Ein Kontrollausdruck setzt die geänderten Daten unter die entsprechenden ursprünglichen, sodaß auch eine Kontrolle der Korrektur rasch und sicher möglich ist.

Die endgültigen Daten können auf mehreren Wegen auf den IBM-Computer gelangen. Bei kleineren Datenmengen ist die direkte Übertragung mittels eines 300 Baud Modems möglich, das an ein serielles Interface des PDP11 angeschlossen wird. Hierbei simuliert das Programm MODEM einen Benützer, der die Daten unter Kontrolle des auf der IBM verfügbaren QED (Quick Editor) "eintippt". Zweckmäßigerweise wird dabei noch das Auswerteprogramm hinzugefügt, das im Falle von SPSS (Statistical Package for the Social Sciences) mittels des äußerst komfortablen Video Editors VITE (1) aus der Parameterliste und vorhandenen Modulen generiert wird.

Die Daten können auch auf Magnetband geschrieben werden, wobei genau definierte Regeln eingehalten werden müssen (2). Normalerweise wird das SPSS-Programm, kombiniert mit Steueranweisungen, durch das Programm CDCREMOTE und das PDP11-FILEX im Interchange-Format auf eine Diskette geschrieben, welche über den HASP-Workstation Satelliten zur IBM übertragen wird.

Auf der IBM Seite existieren nur zwei Programme permanent: Das Fortran Programm TAPETODISK, das alle PDP11 Datenbänder liest, von ihnen ein Disk-File und einen Kontrollausdruck des File-Anfangs und -Endes erstellt, sowie ein SPSS-Preprozessor als Teil der SPSS System-Prozedur. Dieser wurde geschrieben, um die folgenden drei zu SPSS zusätzlichen Anweisungen innerhalb eines SPSS-Programms zu ermöglichen:

	17. Spalte: Kommentar
BEGIN_TEXT	FOLGEND: INAKTIVE BEFEHLE
COMMENT	
COMMENT	FOLGEND: BEISPIEL SPSS
TASK NAME	ALLGEMEINE BESCHREIBUNG
CONDESCRIPTIVE	X1 TO X5
STATISTICS	ALL
COMMENT	
TASK NAME	KOLMOGOROV-SMIRNOV TEST
NPAR TESTS	K-S (NORMAL)= X1 TO X5
COMMENT	
END_TEXT	ENDE INAKTIVE BEFEHLE
FULL_RUN	FOLGENDE BEFEHLE AKTIV

Die Anweisungen BEGIN TEXT / END TEXT, die geschachtelt werden können, verhindern die Übergabe der zwischen ihnen liegenden SPSS Anweisungen an den SPSS-Compiler. FULL RUN bewirkt die Übergabe aller folgenden Anweisungen an SPSS. Der Preprozessor erstellt immer einen vollständigen Programmausdruck mit Kennzeichnung der aktiven bzw. nicht aktiven SPSS-Anweisungen. Daher erfolgt die chronologische Dokumentation des Auswerteverdegangs eines Projektes automatisch, wenn alle zusätzlich durchgeführten Rechnungen am derzeitigen Ende des Programms angefügt werden.

Diese Bearbeitung geschieht ebenso wie das Austesten des SPSS-Programms mit einer Datenstation und einem 300 Baud Akkustik-Modem per Telefon. In Verbindung mit dem QED kann man mit dem ADM31 im Zeilenmode praktisch wie mit einem Video-Editor arbeiten. Die SPSS-Resultate gehen an den Drucker der Satellitenstation.

Nach Abschluß eines Projekts wird das entstandene SPSS Programm als "Punch" auf die HASP - Bandstation geschickt, mit CDCREMOTE in den PDP11 eingelesen und auf einer Diskette zusammen mit den Daten gespeichert. Dieser Gesamt-Backup kann zur Sicherheit noch auf eine ZE 601 Kassette kopiert werden.

Das geschilderte Vorgehen gestattet es, nicht nur auf dem IBM Computer keinerlei Backup unterhalten zu müssen, sondern auch die meisten Tätigkeiten in den gewohnten PDP11-Bereich zu verlegen. Da ferner viele der beschriebenen Schritte in PDP11-Kommandofolgen zusammengefasst sind, ist der Aufwand des Großrechners auch für kleinere Projekte tragbar geworden, wie sich im praktischen Einsatz gezeigt hat.

(1) P. Wegmann, VITE, A Video Text Editor and Corrector for PDP11, Hybridrechenzentrum der ETH Zürich, in Publikation

(2) H. P. Meles Datentransport zwischen Klein- und Großrechner RZU Aktuell Nr. 21., Februar 1978, Universität Zürich